A Comprehensive Evaluation of Methods for Identifying Differentially Methylation Regions in Epigenome-wide Association Studies

Saurav Mallik1, Zhen Gao1, Gabriel J Odam1, Steven Xi Chen1,\*, and Lily Wang1,\*

1Division of Biostatistics, Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, 33136, USA

\*Contact Author: [lily.wang@med.miami.edu](mailto:lily.wang@med.miami.edu), [steven.chen@med.miami.edu](mailto:steven.chen@med.miami.edu)



## **Abstract**

Recent findings show that a number of tools have been developed for the analysis of Differentially Methylation Regions (DMRs). A very limited number of previous studies has been performed regarding the comparison of the methods for predefined regions. However, there is a lack of systematic evaluation of methods that analyze user defined regions. Hence, in this article, we conducted a comprehensive evaluation of the most popular software tools for user defined DMR analysis, including DMRcate, ProbeLasso, Bumphunting and comb-p. In addition to evaluating power, precision and type I error rate, we also compared several additional characteristics of the analysis results by these different methods, including the size of the DMRs, the amount of correlations between nearby CpGs within the identified DMRs, and overlap between the methods.

Key words: DMR finding tools, A-clustering, DMR co-methylation, DMR overlap, DMR size, Evaluation metrics.



**Introduction**

DNA methylation is one of the most studied epigenetic mechanisms, which are stable heritable traits that can not be explained by DNA sequences. The most widely characterized DNA methylation process is the addition of the methyl group at the 5-carbon of the cytosine ring, which results in 5-methylcytosine (5-mC). When located in gene promoter, DNA methylation typically acts to repress gene transcription.

The gold standard for measuring methylation status is the whole-genome bisulfite sequencing (WGBS). However, the high cost of WGBS has limited its use in large epidemiology studies. Currently, most of the Epigenome-wide Association Studies conducted utilized the array based technologies, which provides an economical, high throughput and comprehensive alternative. For example, the HumanMethylation450 BeadChip (Infinium) methylation microarray (450K) (ref) targets 485,577 cytosine positions in the human genome, distributed across gene promoters (within 200 bp or 1500 bp upstream of transcription start sites, exon 1 and in 5´UTRs), gene bodies, 3’UTRs and intergenic regions (Fig 1A). Alternatively, these positions can also be classified by their relation to CpG islands, which are regions in the genome where there are more CG dinucleotides than expected by chance. The CpG positions can be classified into CpG islands (CGIs), shores (2 kb region flanking CGIs), and shelves (2 kb region flanking shores) (Fig 1B). More recently, the MethylationEPIC BeadChip (Infinium) microarray (850K) was developed (ref). This newer array includes more than 90% of the 450K array probes, as well as an additional 333,265 probes targeting sites in regulatory regions identified by the ENCODE (ref) and FANTOM5 (ref) projects recently.

Contiguous regions in the epi-genome with differentially methylated cytosines have been shown to be associated with various diseases (ref). For example, hypermethylation of promoter regions of several candidate genes were shown to be important in neoplastic development and contributor in colon cancer carcinogenesis (ref – cimp papers). In neurodevelopmental disorders, Ladd-Acosta et al. identified and replicated three genomic regions with significant DNAm changes in postmortem brain tissue from ASD patients (ref). Similarly, Liu et al. identified two clusters within the major histocompatibility complex (MHC) region whose differential methylation potentially mediates genetic risk for rheumatoid arthritis (ref).

A number of tools have been developed for the analysis of DMRs (ref). Methods for DMR identification can be classified into supervised and unsupervised methods. The unsupervised methods first groups the CpG probes into genomic regions such as CpG islands, CpG shores, TSS200 using array annotation information, and then test each genomic region for association with phenotype. In contrast, in supervised methods, a p-value or corresponding t-statistic is computed at each CpG first, then the regions in the genome with consecutive small p-values or t-statistics are identified based on user specified criteria such as the minimum number of CpGs for the region. Two previous studies (Klein, Chen) compared unsupervised DMR analysis methods that tests for predefined genomic regions. However, there is a lack of systematic evaluation of supervised DMR identification methods.

In this study, we conducted a comprehensive evaluation of the most popular software tools for supervised DMR analysis, including *DMRcate, ProbeLasso, Bumphunting* and *comb-p*. In addition to evaluating power, precision and type I error rate, we also compared several additional characteristics of the analysis results by these different methods, including the size of the DMRs, the amount of correlations between nearby CpGs within the identified DMRs, and overlap between the methods. We discuss details of the simulation study scheme in “Methods” section and we describe the results of the simulation study and a real dataset in “Results” section. In the last section, we provide a brief discussion on our main findings and also highlight future directions in this research area.

**Methods**

Figure 2 shows the workflow of our simulation study, which involves several steps:

1. First, we obtained a publicly available methylation dataset of 14 healthy controls with similar ages.
2. Next, we performed adjacent site clustering (ref, see details below) to obtain 3063 clusters of adjacent CpG probes (A-clusters).
3. Simulate differentially methylated clusters of CpGs, by randomly selecting 500 A-clusters and adding treatment effects to the simulation datasets. This process was repeated for 5 times for effect sizes .
4. Apply four DMR finding methods (DMRcate, bumphunter, probe-lasso, comb-p) to the simulation datasets generated in step 3.
5. Compare results of these methods on precision, power, type I error rate, size of the DMRs, co-methylation level of CpGs within the DMRs, execution time as well as overlap of DMRs identified by different methods.

We discuss these steps in details next.

**Dataset**

To preserve correlation patterns in real datasets, we generated simulation data by using a real dataset as input. The dataset GSE41169 from Horvath et al. (2012) (ref) included DNA methylation profiles of whole blood samples of 62 schizophrenia patients and 33 healthy controls from the Dutch population. The Illumina Infinium 450k Human DNA methylation Beadchip v1.2 was used to measure the methylation status of 485,577 CpGs. For our study, we selected a total of 14 samples that satisfied two conditions, (1) these are all healthy male samples, and (2) the age-ranges of the patients related to the samples are between 20 and 30. The sample IDs of the 14 selected samples are GSM1009744, GSM1009748, GSM1009666, GSM1009667, GSM1009668, GSM1009688, GSM1009695, GSM1009746, GSM1009742, GSM1009745, GSM1009743, GSM1009681, GSM1009892 and GSM1009893

**Simulating clusters of CpGs with differential methylation**

Before clustering analysis, we used DMRcate function rmSNPandCH to remove CpGs that are close to SNPs, cross-hybridizing and located on sex chromosomes (ref- DMRcate, original paper). We also removed those CpGs with little variations across all the samples, i.e. those CpGs with beta values < 0.05 or beta values > 0.95 in all samples.

Next we used Adjacent Site Clustering (A-clustering) to group neighboring CpGs sites that are correlated with each other into clusters. We applied A-clustering to the 14 samples selected above to obtain a total of 3,063 clusters, each consisting of at least 5 adjacent CpGs. The parameters we used are assign.to.clusters(betas = beta.value, dist.thresh = 0.5, bp.merge = 200, dist.type="spearman", method="complete") which corresponded to merging two CpGs with Spearman correlation greater than 0.5 and are within 200 bp into a cluster. Fig 2 shows an example of an A-cluster with 5 CpGs.

The 14 methylation samples were randomly divided into two groups. Differential methylation of a small subset (i.e. 500) of the clusters were simulated by adding a small treatment effect () to beta values in the group with higher average beta value. On the methylation arrays, beta values are computed based on the ratios of the methylated signal intensity to the sum of both methylated and unmethylated signals after background subtraction, they range from 0 (completely unmethylated ) to 1 (fully methylated).

**Methods for identifying DMRs**

Table 1 list the parameter settings used for the four DMR finding methods. We discuss each of the methods in details next.

**DMRcate**

DMRcate method is implemented in Bioconductor package DMRcate (ref). The DMRcate method first fits a linear model at each CpG using the empirical Bayesian methodology from *limma* R package. In our study, this model included methylation M value as the outcome variable and group status as independent variable. M values are logit transformation of beta values, that is M = log (beta/1-beta), it has been shown to have better statistical properties such as homoscedasticity (ref) in methylation data analysis. The statistic *Y = t2* is then calculated for each position, where *t* is the t-statistics from linear model corresponding to group effect. In the second step, DMRcate applies kernel smoothing using the Gaussian smoother. P-values for each position is then computed by moment matching using the method of Sattererthwaite (ref). The CpG sites with multiple comparison corrected (via the method of Benjamini–Hochberg (ref)) are then selected as significant CpGs. Regions for DMRs are identified by collapsing contiguous significant CpGs that are at most nucleotides from each other. The p-value for DMR is computed using Stouffer’s method ([Hoffman, 1965](#_ENREF_2)).

**bumphunter**

Bumphunter is implemented in the Bioconductor packages bumphunter and minfi (ref). In the bumphunting method, first a linear regression model M value ~ Group is applied to model differential methylation between case and control groups at each CpG site. Candidate regions (bumps) are identified to be clusters of consecutive probes for which all the t-statistics exceed a user defined threshold (argument cutoff in bumphunter function). Permutation test, which permutes sample labels to create null distribution of candidate regions, are then conducted to estimate statistical significance of the candidate regions. Note that in the identification of regions, spatial correlation structure were used to model correlations of methylation levels between neighboring CpGs.

**Probe Lasso**

Probe Lasso is implemented in the Bioconductor package ChAMP (ref). In Probe Lasso method, first a linear regression model Beta value ~ Group is applied to model differential methylation between case and control groups at each CpG site. Next, because the probe spacing on the methylation arrays are not uniform (e.g. probes located in promoter regions are more densely spaced, while those located in intergenic regions are more spread out), Probe Lasso defines flexible boundaries around each probe depending on the type of genomic feature the probe is located in (e.g. TSS200, 3’UTR). Much like a real lasso, Probe Lasso are then “thrown” around each probe with the dynamic boundaries, centered at the target probe. A region around the target probe is selected if the number of significant probes captured within the probe-lasso boundary is higher than the user-specified threshold (argument minProbes in champ.DMR function). For each region, Probe Lasso then computes a correlation matrix of normalized beta values within each region, and then uses Stouffer’s method to compute p-value for the region, by weighting individual probes by the inverse sum of its squared correlation coefficient in the correlation matrix.

**comb-p**

Comb-p is a command-line tool and a python library ([Pedersen, et al., 2012](#_ENREF_5)). In contrast to the three methods described above, it does not support calculation of p-values for individual CpG. Instead, the input of comb-p is a BED file with p-values and chromosome locations of the CpG sites. Comb-p then computes correlations at varying distance lags (auto-correlation or ACF), which are used to compute corrected p-value at each CpG site using Stouffer-Liptak-Kechris correction (refs). The corrected p-value at a CpG site will be lower than the original p-value if the neighboring CpG sites also have low p-values. On the other hand, the corrected p-value at a CpG site will remain insignificant if neighboring p-values are also high. The false discovery rate (ref) is then calculated at each CpG site and a peak-finding algorithm is then used to find regions enriched with small p-values. Once the regions are identified, the final p-value for each region is then computed based on the Srouffer-Liptak correction.

< insert Table 1 about here >

Results

**Power, precision and type I error**

Figure

- Power: Power is defined as the number of true positive DMRs divided by the total of the number of true positive and false negative DMRs (i.e., power=TP/(TP+FN)).

- Precision: Precision is defined as as the number of true positive DMRs divided by the total of the number of true positive and false positive DMRs (i.e., precision=TP/(TP+FP)).

- DMRsize: DMRsize is stated as the number of CpGs in any resultant DMR obtained by any DMR finding method.

- DMR co-methylation: DMR co-methylation is here defined as the average of Spearman correlation coefficient between the participating CpGs in pair-wise manner.

- Time used: The elapsed time is also measured in Second. The configuration of the used computer is: ###.

- DMR-overlap: DMR-overlap is here stated as the number of common significant resultant DMRs (i.e., the DMRs having p-value less than 0.05 and DMRsize is greater than or equal to 5) obtained by the DMR finding methods pairwise. Of note, it may happen that one DMR determined from any method can overlap with multiple DMRs identified from other method, or vice versa.

- Evaluation metrics: Three evaluation metrics (viz., true positives or TPs, false positives of FPs, and false negatives or FNs) have been utilized for our analysis. The definition of these metrics for our study and the other related information regarding these metrics are demonstrated in Table 2.

**Evaluation criteria**

- Power: Power is defined as the number of true positive DMRs divided by the total of the number of true positive and false negative DMRs (i.e., power=TP/(TP+FN)).

- Precision: Precision is defined as as the number of true positive DMRs divided by the total of the number of true positive and false positive DMRs (i.e., precision=TP/(TP+FP)).

- DMRsize: DMRsize is stated as the number of CpGs in any resultant DMR obtained by any DMR finding method.

- DMR co-methylation: DMR co-methylation is here defined as the average of Spearman correlation coefficient between the participating CpGs in pair-wise manner.

- Time used: The elapsed time is also measured in Second. The configuration of the used computer is: ###.

- DMR-overlap: DMR-overlap is here stated as the number of common significant resultant DMRs (i.e., the DMRs having p-value less than 0.05 and DMRsize is greater than or equal to 5) obtained by the DMR finding methods pairwise. Of note, it may happen that one DMR determined from any method can overlap with multiple DMRs identified from other method, or vice versa.

- Evaluation metrics: Three evaluation metrics (viz., true positives or TPs, false positives of FPs, and false negatives or FNs) have been utilized for our analysis. The definition of these metrics for our study and the other related information regarding these metrics are demonstrated in Table 2.

Table 2: The definition of the initial three factors (i.e., true positive (TP), false positive (FP), and false negative (FN)) of the confusion matrix in our simulation study.

Results

**Results of Simulation Study**

In the experiment, we observe the comparative performance of the four DMR finding methods through the alternation of *µ*.

Power

For Smaller µ value (0-0.05)

In the case of smaller µ, only a few TPs had been generated in any resultant DMR finding method for any repetition (run). Bumphunter produced highest number of TPs rather than the other three methods, whereas Probe-lasso generated lowest number of TPs among all. Overall, Bumphunter provided highest average power in this scenario (viz., 0.2023(±0.02) for miu=0.025, and 0.5002(±0.02) for mu=0.05), whereas comp-b yielded second highest power (viz., 0.0852(±0.01) for miu=0.025, and 0.3104(±0.02) for mu=0.05). But, the performance of the remaining two methods (DMRcate and Probe-lasso) were poor.

Overall, no method works well when µ value is low. In addition, it can be said that Bumphunter performs better than the others for smaller value of µ.

For Medium µ value (0.1-0.15)

In the case of medium µ, the number of TPs were increased than these in the smaller µ for the four DMR identification methods. In this case also, Bumphunter was the best performer in producing highest power (viz., 0.7451(±0.03) for miu=0.1, and 0.7805(±0.02) for mu=0.15) among the all methods, whereas DMRcate and Probe-lasso generated lower power (viz., 0.3832(±0.01) and 0.3892(±0.02) respectively for mu=0.1, and 0.4744(±0.02) and 0.4628(±0.02), respectively for mu=0.15) among the all. The remaining method Comb-p second higher average power generator.

Overall, the performance of all the methods improve significantly for medium µ value rather than these for low µ value. In this case also, Bumphunter is the best performer.

For Higher µ value (0.2-0.4)

For higher µ, the number of TPs were increased than these in the medium µ for the four methods. Alike the previous cases, Bumphunter produced highest power (viz., 0.9918(±0) for miu=0.2, 0.9923(±0) for miu=0.3, and 0.9926(±0) for mu=0.4) among the all methods, whereas Probe-lasso generated lower power among the all. Comb-p was the second highest average power generator.

Overall, for higher µ value, all methods perform well. Moreover, alike in the previous cases, Bumphunter is the best performer here.

See Figure 2 and Figure 3 for details. For details about the overall comparative performance of the four DMR finding methods, see Table 3.

Precision

For Smaller µ value (0-0.05)

On the other hand, Comb-p and DMRcate were the best performers in terms of the identified precision (viz., 0.9963(±0.01) and 1(±0) respectively for mu=0.025, and 0.9963(±0) and 0.9831(±0.01), respectively for mu=0.05). Of note, the performance of Comb-p was consistent for smaller µ, wheres Bumphunter produced lowest precision in this case (0.7058(±0.02) for miu=0.025, and 0.8722(±0.01) for mu=0.05) since it generated a lot of FPs rather than these of the other methods.

However, overall, only two methods (Comb-p and DMRcate) work well for µ<0.05.

For Medium µ value (0.1-0.15)

On the other hand, Comb-p was the best performer in terms of the identified precision (viz., 0.9917(±0) for mu=0.1, and 0.9924(±0) for mu=0.15). Specially, Comb-p performed consistently for the medium µ. DMRcate and Bumphunter yielded lowest precision (as well as inconsistent) in this case.

Overall, all the methods more or less perform well for the medium µ value.

For Higher µ value (0.2-0.4)

On the other hand, Comb-p was the best performer in terms of the precision (viz., 0.8199(±0.02) for miu=0.2, 0.8687(±0.03) for miu=0.3, and 0.9099(±0.03) for mu=0.4). Specially, Comb-p and probe-lasso performed consistently for the higher µ. DMRcate generated lowest precision in this case.

Overall, all the methods work well for the higher µ value.

Time

In addition, …….  **###time (need zen’s time calculation on comb-p)**

DMR Overlap

For Smaller µ value (0-0.05)

We also identified the overlap between the four DMR finding methods for smaller µ. Of note, since a DMR of any method can overlap with one or multiple DMRs of the other method in partially or completely, there exists some one-to-many (or, many to one) mapping in the results obtained from the methods. For µ=0.025 and repetition=1, we obtained 8 common DMRs from DMRcate method and 9 common DMRs from Bumphunter method during the overlap operation between DMRcate and Bumphunter method. Similarly, we determined 22 common DMRs from Bumphunter method and 19 common DMRs from Comb-p method during the overlap operation between Bumphunter and Comb-p methods. See Table 4 for details. For other µ values, see supplementary file ST1.

For Medium µ value (0.1-0.15)

In case of the overlap operation, for µ=0.15 and repetition=1, we identified 220 common DMRs from DMRcate method and 354 common DMRs from Bumphunter method during the overlap operation between DMRcate and Bumphunter method. Similarly, we obtained 458 common DMRs from Bumphunter method and 301 common DMRs from Comb-p method during the overlap operation between Bumphunter and Comb-p methods, whereas 198 common DMRs from Probe-lasso method and 200 common DMRs from Comb-p method during the overlap operation between Probe-lasso and Comb-p methods. See Table 4 for details. For other µ values, see supplementary file ST1.

For Higher µ value (0.2-0.4)

During the overlap operation, for µ=0.4 and repetition=1, we identified 374 common DMRs from DMRcate method and 605 common DMRs from Bumphunter method during the overlap operation between DMRcate and Bumphunter method. Similarly, we obtained 575 common DMRs from Bumphunter method and 360 common DMRs from Comb-p method during the overlap operation between Bumphunter and Comb-p methods, whereas 221 common DMRs from Probe-lasso method and 222 common DMRs from Comb-p method during the overlap operation between Probe-lasso and Comb-p methods. See Table 4 for details. For other µ values, see supplementary file ST1.

Co-methylation

For Smaller µ value (0-0.05)

Furthermore, we computed average Spearman’s correlation of the participating CpGs belonging to the top 5 DMRs. Bumphunter generated highest correlation in all these cases, but bumphunter suffered from a lot of outliers. See Figure 4(a), and supplementary file SF1(d) for details.

For Medium µ value (0.1-0.15)

For medium µ, Bumphunter and Probe-lasso perform well in terms of the average Spearman’s correlation measures. But, Bumphunter is still the best performer in maximum repetitions. See Figure 4(b), and supplementary files SF1(a) and SF1(e) for details.

For Higher µ value (0.2-0.4)

For higher µ, Bumphunter and Comb-p work well in terms of the average Spearman’s correlation measures. But, Bumphunter is still the best performer in maximum repetitions. See Figure 4(c), and supplementary files SF1(b), SF1(c) and SF1(f) for details.

Sizes of DMRs

For higher µ value (e.g., µ =0.4), the average number of CpGs in the top 5 DMRs obtained by DMRcate is maximum (~10), whereas the same number obtained by Bumphunter is minimum (~5). See Figure 5 for details.

For other µ values, see supplementary files SF2(a)-SF2(f).

**Results of Real Data**

**Discussion and Conclusion**

Identifying differentially methylation regions is a latest topic of interest. As per recent findings, a very few previous studies has been conducted to compare the DMR finding methods for predefined regions. Moreover, there is a lack of systematic evaluation of methods that analyze user defined regions. Therefore, in this article, we conducted a comprehensive evaluation of the most popular software tools for user defined DMR analysis, including DMRcate, ProbeLasso, Bumphunting and Comb-p. First of all, we applied A-clustering to the 14 healthy control samples with similar ages to obtain a total of 3,063 clusters, each consisting of at least 5 adjacent CpGs. The above samples were then randomly divided into two groups, and differential methylation of a small subset of the clusters were simulated by adding a small number to beta values in one of the groups that contains higher mean beta-value. Next, we ran the aforementioned four methods with default parameters to identify DMRs. Of note, for evaluation, we utilized three metrics (viz., TP, FP and FN), and several criteria such as power, precision, time, DMR overlap, DMR co-methylation, and DMRsize. From the outcome, it has been observed that for the power estimation, Bumphunter is best, where for the precision, Comb-p and DMRcate performed best. For co-methylation, Bumphunter is best. In case of DMRsize, DMRcate generated maximum number of average CpGs determined from the specified number of top resultant DMRs, whereas Bumphunter provided minimum number of average CpGs determined from the specified number of top resultant DMRs. Overall, Bumphunter works more or less best among all the methods in terms of all evaluation criteria, but it took a long elapsed time rather than the others. On the other hand, Probe-lasso was probably faster than the others. Specially, all the methods performed well for medium of higher µ value, but for small µ value, no method works well.

In addition, there are several issues raised over these methods. Firstly, the number of choices for choosing parameters are large for these methods, and these parameters are not properly described how to set them. Although in this study, we used default parameters, it is hoped that the future developers clarify different parameter setting for users. Secondly, for small µ value, the performance of these methods is questionable. Thirdly, the regions might not be precisely specified for Bumphunter. Finally, our comparative study provides the critical review of the latest well-known DMR finding methods through which the future users understand the advantages and shortcomings of the underlying methods and can develop new method to resolve the aforementioned issues.

We hope our results based on both simulation study and real methylation datasets will be helpful to investigators for selecting the most appropriate methods for their studies.

**Legends for Figures and Tables**

Table 3: True Positives (TP), False Positives (FP), False Negatives (FN), Power, Precision and Elapsed Time (in Second) for the four DMR finding methods in the simulation study.

Table 4: Overlap between the four DMR finding methods for low (0.025), medium (0.15) and large (0.4) *µ* for the four DMR finding methods in the simulation study.

Figure 1: Overall workflow of the analysis.

Figure 2: Power comparison of the four DMR finding methods in the simulation study.

Figure 3: Precision comparison of the four DMR finding methods in the simulation study.

Figure 4: Average pairwise correlation for top five DMRs obtained from the four DMR finding methods in the simulation study.

Figure 5: nCpGs for significant DMRs obtained from the four DMR finding methods in the simulation study.

**References**

Butcher, L.M. and Beck, S. (2015) Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data, *Methods*, **72**, 21-28.

Hoffman, J. (1965) The American Soldier, Vol 1, Adjustment during Army Life, Vol 2, Combat and Its Aftermath - Stouffer, *Contemp Psychol*, **10**, 505-505.

Kechris, K.J., Biehs, B. and Kornberg, T.B. (2010) Generalizing Moving Averages for Tiling Arrays Using Combined P-Value Statistics, *Stat Appl Genet Mol*, **9**.

Morris, T.J.*, et al.* (2014) ChAMP: 450k Chip Analysis Methylation Pipeline, *Bioinformatics*, **30**, 428-430.

Pedersen, B.S.*, et al.* (2012) Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values, *Bioinformatics*, **28**, 2986-2988.

Rindskopf, D. (1997) An introduction to the bootstrap - Efron,B, Tibshirani,RJ, *J Educ Behav Stat*, **22**, 245-245.

TEXT for review paper

All the DMR analysis methods follows three steps (1) fits linear model to each CpGs (2) identify candidate regions in the genome with consecutive CpGs with small p-values (3) compute p-values for the candidate regions. We discuss in details for each of these steps for DMRcate, bumphunter, probeLasso and comb-p in the following.

Summary of the methods

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Single CpG | smoothing | Covariate allowed? | P-value for DMR | Software implementation |
| bumphunter | Linear model | Optional loess | Yes | Permutation of sample labels or bootstrap | bumphunter  minfi |
| DMRcate | limma | Gaussian smoothing | Yes | Stouffer’s method | DMRcate |
| ProbeLasso |  |  |  |  |  |

Text

**DMRcate**

For the the DMRcate method, first a linear model with methylation M value as the outcome variable, group status and any other covariate variables such as age or batch effects is applied to data for each individual CpG by using functions from the limma R package. *limma* is one of the most popular statistical tool for assessing differential gene expression expressions. In the application to methylation study, because *limma* shrinks variances for each CpG toward the global variance estimated from all CpGs using the empirical Bayes method, results from limma are more stable for studies with small sample sizes.

The statistic Y = t2 is calculated for each position, where t is the t-statistics from linear model corresponding to group effect. In the second step, DMRcate applies kernel smoothing using the Gaussian smoother. P-values for each position is then computed by moment matching using the method of Sattererthwaite (ref). The CpG sites with with multiple comparison corrected (via method of BH) are selected as significant CpGs.

Regions for DMRs are identified by collapsing contiguous significant CpGs that are at most lamda nucleotides from each other. The p-value for DMR is computed using Stouffer’s method ([Hoffman, 1965](#_ENREF_2)).

**Bumphunter**

In the bumphunting method, first a linear regression model M value ~ Group is applied to model differential methylation between case and control groups at each CpG site. Here M value is a logit transformation of the beta values, i.e. M value = log2 (beta value / (1- beta value). It has been shown that in the analysis of methylation data, M values have better statistical properties such as homoscedasticity (ref). Next, an optional smoothing step is involved where the loess curve fitting is applied to estimated t-statistic corresponding to the regression coefficients. Candidate regions (bumps) are identified to be clusters of consecutive probes for which all the t-statistics exceed a user defined threshold (argument xx in bumphunter function). Permutation test, which permutes sample labels to create null distribution of candidate regions, are then conducted to estimate statistical significance of the candidate regions.

The bumphunter function is implemented in *minfi* R package. Note that in the identification of regions, spatial correlation structure were used to model correlations of methylation levels between neighboring CpGs. When design of the study involves covariate variables, such as age or batch effect, bootstrap ([Rindskopf, 1997](#_ENREF_7)) option can be used to estimated statistical significance of the bumps.

####Ref Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York, NY: Chapman and Hall, 1993, p. 436.

The bumphunter function is implemented in *minfi* R package. When design of the study involves covariate variables, such as age or batch effect, bootstrap ([Rindskopf, 1997](#_ENREF_6)) option can be used to estimated statistical significance of the bumps.